PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04L 12/56, H04Q 11/04 — A1

(11) International Publication Number: WO 97/14240

(43) International Publication Date: 17 April 1997 (17.04.97)

(21) International Application Number: PCT/CA96/00681

(22) International Filing Date: 11 October 1996 (11.10.96)

(30) Priority Data:
9520807.0 11 October 1995 (11.10.95) GB

(71) Applicant *(for all designated States except US)*: NEWBRIDGE NETWORKS CORPORATION [CA/CA]; 600 March Road, P.O. Box 13600, Kanata, Ontario K2K 2E6 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GIROUX, Natalie [CA/CA]; 8 de la Comète, Hull, Québec J9A 2Y6 (CA). LIAO, Raymond, Rui-Feng [CA/CA]; Unit 903, 5785 Yonge Street, North York, Ontario M2M 4J2 (CA). AISSAOUI, Mustapha [CA/CA]; 30 Eleanore Drive #1210, Nepean, Ontario K2E 7E5 (CA).

(74) Agent: MITCHELL, Richard, J.; Marks & Clerk, P.O. Box 957, Station B, Ottawa, Ontario K1P 5S7 (CA).

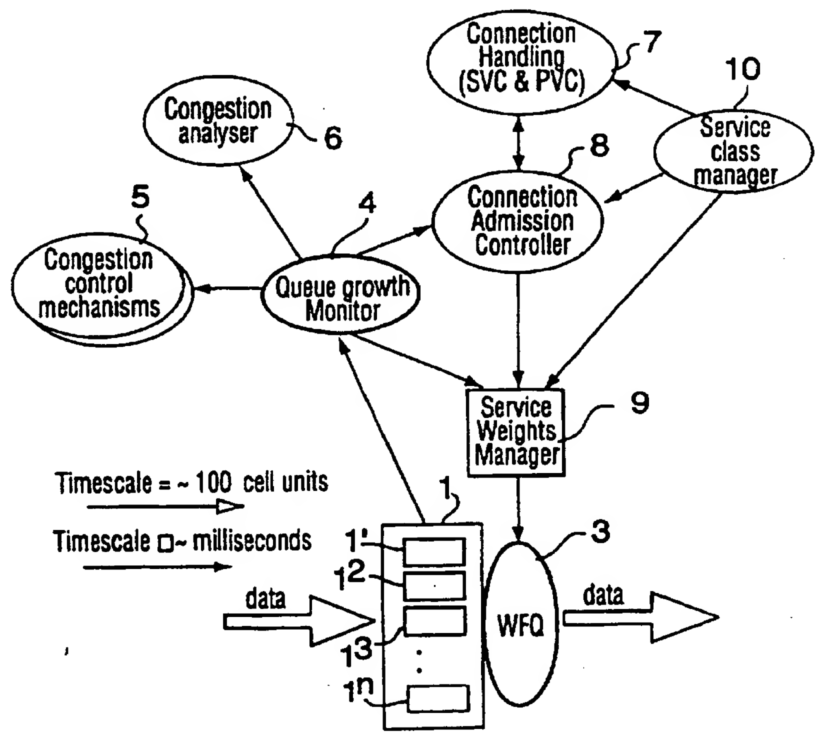
(81) Designated States: AU, CA, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: FAIR QUEUE SERVICING USING DYNAMIC WEIGHTS (DWFQ)

(57) Abstract

In a method of fair queue servicing at a queuing point in a multi-service class packet switched network, incoming packets are received in buffers and outgoing packets are scheduled by a weighted fair queue scheduler. Real-time information of buffer usage along with the minimum bandwidth requirement is used to dynamically modify the weights of the weighted fair queue scheduler.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

Fair Queue Servicing using dynamic weights (DWFQ)

This invention relates to the field of telecommunications, and more particularly to a method of fair queue servicing in asynchronous data networks, such as Asynchronous Transfer Mode (ATM) networks or more generally any packet switched network that supports more than one class of service.

The use of ATM by a continually increasing number of applications is driving a requirement to increase the number of service classes and to allow more flexibility in the service offerings. To support the application requirements, the ATM Forum is adding new service categories in new releases of ATM specifications. Furthermore, network providers are looking for the flexibility of defining multiple service classes for a given service category. The service classes are differentiated by their Quality-Of-Service requirements (QoS). The QoS requirements are configurable in accordance with a bi-dimensional matrix describing loss and delay. The delay jitter is another factor which needs to be bounded for some service classes.

Previously, three service categories were supported on an ATM network element, namely constant bit rate (CBR), variable bit rate (VBR) and unspecified bit rate (UBR). The CBR service is the only service that guarantees a bound on delay. It is used for time sensitive data, such as voice and video.

These various services can be supported by traditional exhaustive round-robin queuing among two priority queues. However, this simple technique cannot be used when the number of queues increases beyond two, because of the high potential of starvation for lower priority queues. Furthermore, the exhaustive round robin can only guarantee bounds on delay and delay variation for the highest priority queue. The support of multiple service class in an ATM switching product or multiplexer requires a minimum of one queue per class.

A queue scheduling algorithm, Weighted Fair Queuing (WFQ), has been recently proposed in the literature (see S. Golestani, A self-clocked Fair Queuing scheme for broadband applications. INFOCOM 1994. June 1994).

This scheduling scheme allows any number queues (service classes) to be serviced, while providing fair and work conserving access to bandwidth. One of the key features of WFQ is that the CDV (Cell Delay Variation) is bounded for any service class, as long as it is given a minimum weight.

This proposed scheme can be implemented in ATM products. However, it has not been determined how to set the servicing weights efficiently to take into account the dynamically changing bandwidth requirement of each service class (connection addition/removal, ABR flow control, Early packet Discard).

An object of the invention is to provide a framework that ensures that the weights are set appropriately to guarantee the desired Quality of Service and modified in real-time to ensure that the dynamic allocation of bandwidth across the classes is optimized.

According to the present invention there is provided a method of fair queue servicing at a queuing point in a multi-service class packet switched network, wherein incoming packets are received in buffers and outgoing packets are scheduled by a weighted fair queue scheduler characterized in that real-time information of buffer usage along with the minimum bandwidth requirement is used to dynamically modify the weights of the weighted fair queue scheduler.

Preferably the minimum bandwidth requirement is extracted during connection admission control.

The method is particularly suitable for use in ATM networks.

The DWFQ (Dynamic Weighted Fair Queuing) can be implemented at any queuing point which arbitrates servicing between *n* queues ($n \geq 2$).

The invention also provides a fair queue servicing arrangement in a multi-service class packet switched network, comprising a weighted fair queuing controller, and buffer means for receiving incoming packets in queues, characterized in that further comprises means for monitoring buffer usage for each queue, means for determining the bandwidth requirements of each class of service, and a service weights manager for dynamically modifying the weights of said weighted fair queuing controller means in response to said buffer usage and bandwidth requirements.

Preferably, the means for monitoring buffer usage a queue growth monitor which performs real-time estimation of the queue growth in said buffer means.

The invention will now be described in more detail, by way of example only, with reference to the accompanying drawings, in which:-

Figure 1 is a diagram depicting the high level queuing scheme at an ATM switch;

Figure 2 shows the information provided by the Queue Growth Monitor;

Figure 3 illustrates the data flow between the key components of the system and the action of the Service Weight Manager (SWM); and

Figure 4 describes the process performed by the SWM.

Referring now to Figure 1, ATM cells 2 arrive at buffer 1 and are placed in queues 1^1, 1^2, ... 1^n. From there the cells are passed to a weighted fair queuing unit 3. The buffer 1 is also connected to a queue growth monitor 4, which in turn is connected to congestion control unit 5, congestion analyzer 6, and connection admission controller 8, which in turn is connected to SVC & PVC (Switched Virtual Circuit and Permanent Virtual Circuit) connection handling unit 7, and service class manager 10. Queue growth monitor 4, connection admission controller 8 and service class manager 9 are connected to service weights manager 9, which is connected to weighted fair queuing scheduler 3.

The key element of the Dynamic Weighted Fair Queuing (**DWFQ**) scheme is the service weight manager (SWM) 9, which dynamically modifies the service weights to be used by the WFQ Scheduler 3. It uses real-time information from the service class manager 10, the connection admission controller 8, and the Queue growth monitor 4.

The service class manager 10 configures the service classes. A service class is configured with a given value of delay (CTD - Cell Transfer Delay) and loss (CLR - Cell Ratio Loss) requirements. These parameters represent the maximum nodal delay and loss allowed in order to meet the end-to-end QoS requirements of the connection. The service classes are mapped into a priority table as exemplified in Table 1. The priority table is used later by the service weight manager to allocate remaining bandwidth. The priority table is updated when a service class definition is modified. The service class manager

also dictates which traffic descriptors are used to compute the minimum bandwidth required by a connection of a given class.

	CTD		
CLR	100 µs	500 µs	None
10^{-9}	1	3	6
10^{-7}	2	5	8
10^{-5}	4	7	9

Table 1 • Example of a Queue Service Priority Mapping.

The connection admission controller (CAC) 8 computes the minimum bandwidth required for each service class. The minimum bandwidth is updated each time a connection of a given class is established or disconnected, based on its traffic descriptor.

Table 2 shows a typical example of which traffic descriptors that can be used to compute the minimum bandwidth for each basic service category relative to the queue service rate (SR). The CAC 8 communicates the minimum Weight table to the SWM every time the value of the minimum weights have changed by a factor of ξ.

$Queue_i$	Category	min_Wi
Q_1	CBR	(Σ PCR) / SR
Q_2	RT-VBR	(Σ SCR) / SR
Q_3	NRT-VBR	(Σ SCR) / SR
Q_4	UBR	0

Table 2 • Example of minimum weight table.

The CTD is further taken into account in the target queue size (TQS) table, which is the maximum queue size allowed to limit the CTD. An example of TQS_i computation is shown in Table 3, for typical service categories. A zero TQS indicates that the queue can grow without limitation. This table is computed by the CAC.

$Queue_i$	Category	TQSi
Q_1	CBR	CTD / min_Wi
Q_2	RT-VBR	CTD / min_Wi
Q_3	NRT-VBR	0
Q_4	UBR	0

Table 2 • Example of a target queue size computation.

The Queue Growth Monitor (QGM) 4 performs real-time estimation of the queue growth every $\mathbf{T_S}$ cell slots (sampling interval). The information provided by the Queue Growth monitor 4 to the SWM 9 consists of **ΔQi**, the Queue Growth Rate of output queue i during an interval of duration **Ts, Qi**, the length of output queue i at the sampling time, and **Ai**, the arrival rate during the same interval of time.

The corresponding parameters: queue size Qi, queue growth ΔQi, number of arrivals Ai are collected or each queue by the QGM 4 for each Ts interval. From these parameters, auxiliary parameters such as average arrival rate λi and service rate μi can be derived by the SWM 9:

λ_i : average arrival rate, $\lambda_i = A_i \div T_s$

μ_i : average service rate, $\mu_i = S_i \div T_s$, where $S_i = A_i - \Delta Q_i$ is the number of cells served during T_s.

Figure 2 shows the information provided by the Queue Growth Monitor 4. Using this information for the CAC 8, the service class manager 9 and the queue growth monitor 4, the SWM computes the service eight for each queue i (W_i) to be used during the next sampling interval.

As can be seen in Figure 3, which shows the data flow between the key components of the system and the action of the SWM 9, the queue weights, Wi, are updated using information provided by the Queue Growth monitor 4.

If $\tilde{\lambda}_i$ denotes the arrival rate of cells in queue i in the coming T_s interval, then ideally, the target service rate $\tilde{\mu}_i$ can be calculated as: $(\tilde{\mu}_i - \tilde{\lambda}_i) \cdot T_s = Q_i - TQS_i$. This means at the end of next Ts interval, the queue size Qi will reach the target queue size

TQSi. On the assumption that $\tilde{\lambda}_i$ remains unchanged from λ_i, the service weight $W_i = \tilde{\mu}_i \bullet T_s$ can be approximated as

$$W_i = \tilde{\mu}_i \bullet T_s \cong \lambda_i \bullet T_s + Q_i - TQS_i = A_i + Q_i - TQS_i .$$

However, the assumption on the stable arrival rate may not hold, and also the actual number of serviced cells Si could be less than Wi: therefore a more conservative approach is $\Delta Q_i > 0$, then $Q_i+\Delta Q_i$, the predicted queue size at the end of the next Ts interval, is used to calculate the target service rate and weight. That is:

$$\left(\tilde{\mu}_i - \tilde{\lambda}_i\right) \bullet T_s = Q_i + \Delta Q_i - TQS_i \text{ and}$$

$$W_i \cong \lambda_i \bullet T_s + Q_i + \Delta Q_i - TQS_i = A_i + Q_i + \Delta Q_i - TQS_i$$

The detailed algorithm performed by the service weights manager 9 is shown in Figure 4. The queue size Q_i at the end of each interval T_s, the number of arrivals A_i during the previous interval T_s, and the change in queue size ΔQ_i are input at step 20. Step 21 determines whether the queue growth is positive: if yes, the service weight W_i is conservatively adjusted to bring the queue size to TQSi at step 22; if no, the service weight W_i is adjusted to bring the queue size to TQSi at step 23. The difference ΔW is determined in step 24.

Step 25 determines whether the shared weights pool is empty: if yes, Wi is set to min_Wi in step 26; if no, step 27 determines whether $Ws \geq \Delta W_i$: if yes, step 28 sets $W_i = min\_W_i + \Delta W_i$ and $W_s = W_s + \Delta W_i$; if no, step 29 sets $W_i = min\_W_i + W_s$ and $W_s = 0$.

Step 30 runs through all the W_i in the ordered list L and step 31 updates the weight table used by the Weighted Fair Queuing scheduler 3.

The described technique complies with ITU and ATM Forum standards and can be applied to any switching equipment which supports more than a single service class of service.

Claims:

1. A method of fair queue servicing at a queuing point in a multi-service class packet switched network, wherein incoming packets are received in buffers and outgoing packets are scheduled by a weighted fair queue scheduler characterized in that real-time information of buffer usage along with the minimum bandwidth requirement is used to dynamically modify the weights of the weighted fair queue scheduler.

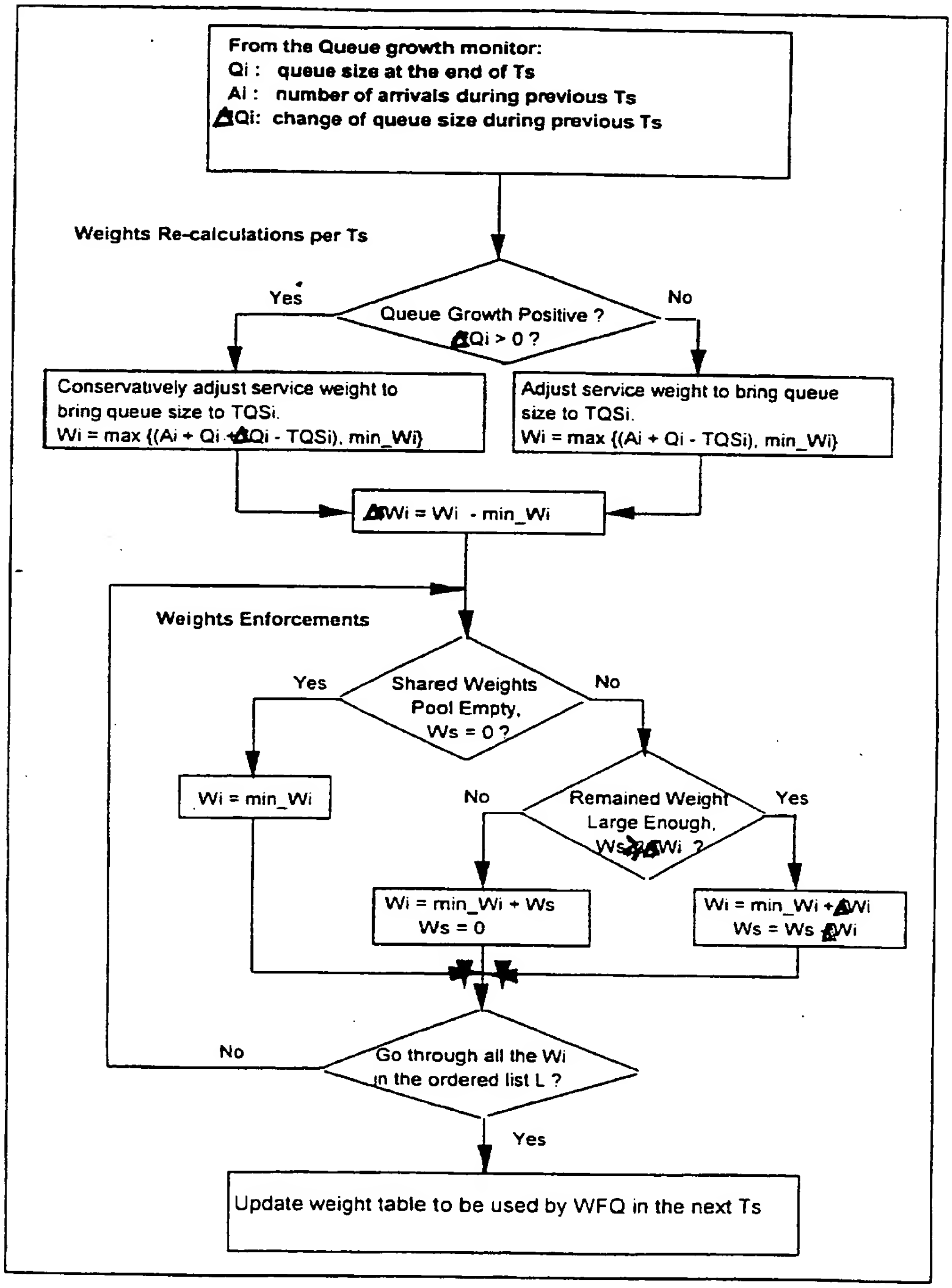
2. A method as claimed in claim 1, the minimum bandwidth requirement is extracted during connection admission control.

3. A method as claimed in claim 1, characterized in that said weights are also modified in accordance with real-time service class information.

4. A method as claimed in claim 1, characterized in that buffer usage is monitored by a queue growth monitor, which performs real-time estimation of the queue growth every sampling interval T_s.

5. A method as claimed in any of claims 1 to 4, wherein said packet switched network is an ATM network and said packets are ATM cells.

6. A method as claimed in any of claims 1 to 5, wherein the weights of the weighted fair queue scheduler are modified in accordance with the following algorithm:



7. A fair queue servicing arrangement in a multi-service class packet switched network, comprising a weighted fair queuing controller, and buffer means for receiving incoming packets in queues, characterized in that further comprises means for monitoring buffer usage for each queue, means for determining the bandwidth requirements of each class of service, and a service weights manager for dynamically modifying the weights of said weighted fair queuing controller means in response to said buffer usage and bandwidth requirements.

8. A fair queue servicing arrangement as claimed in claim 7, characterized in that it further comprises a service class manager which stores the cell transfer delay cell loss ratio requirements for each class of service, and said service weights manager is also responsive to said class of service requirements stored in said service class manager to dynamically modifythe weights of said weighted fair queuing controller means.

9. A fair queue servicing arrangement as claimed in claim 8, characterized in that said means for monitoring buffer usage comprises a queue growth monitor which performs real-time estimation of the queue growth in said buffer means.

10. A fair queue servicing arrangement as claimed in claim 8, characterized in that said queue growth monitor outputs to said service weights manager the queue growth rate of each output queue *i* during a sampling interval, the length of the output queue *i* at sampling time, and the arrival rate of cells A_i during the sampling interval.

11. A fair queue servicing arrangement as claimed in any of claims 8 to 10, characterized in that said network is an ATM network.

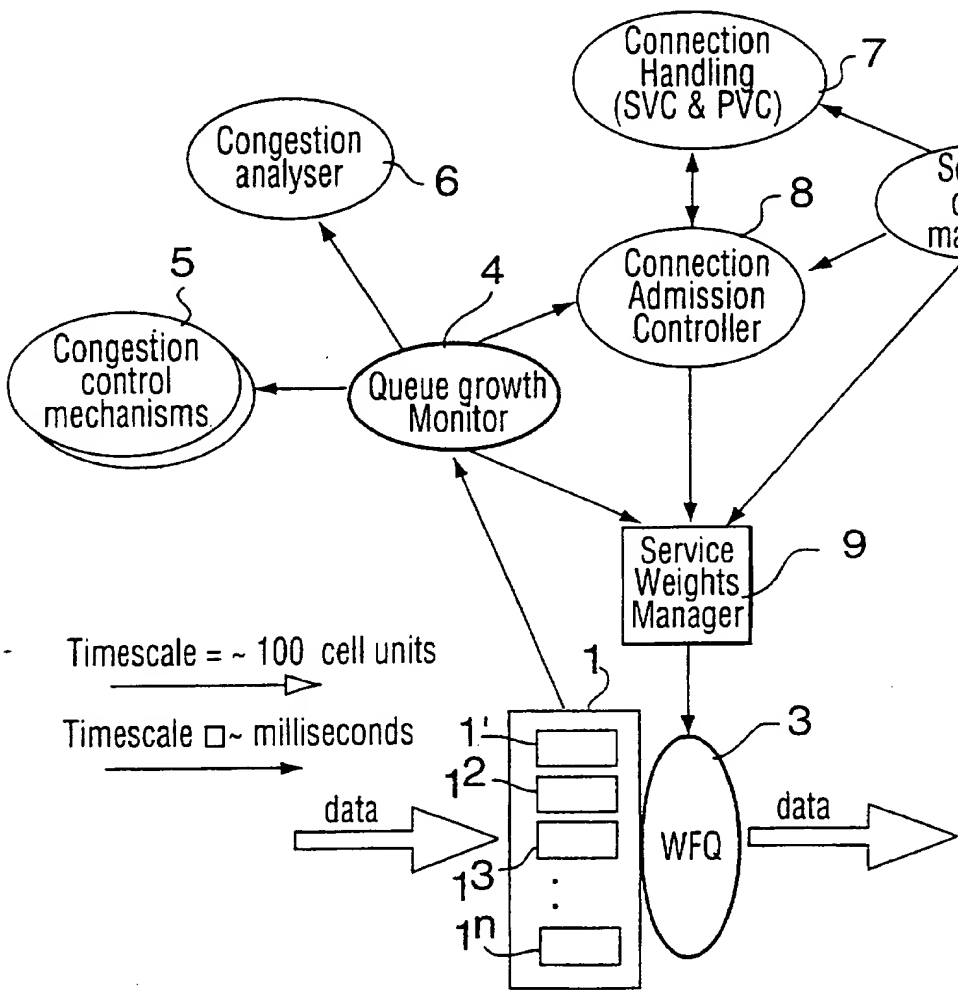

Connection Handling (SVC & PVC)
7
10
Congestion analyser
6
8
Service class manager
5
4
Connection Admission Controller
Congestion control mechanisms
Queue growth Monitor
Service Weights Manager
9
Timescale = ~ 100 cell units
1
Timescale □~ milliseconds
3
1'
12
data
13
WFQ
data
1n

FIG. 1

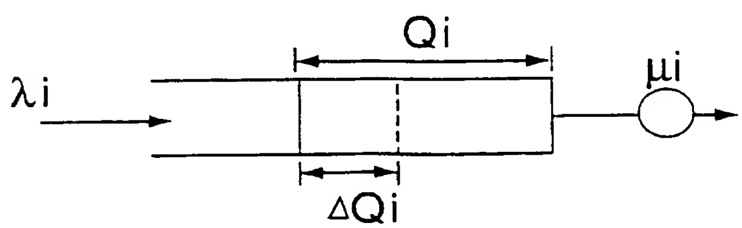

Qi
λi
μi
ΔQi

FIG. 2

Ts: Sampling interval
Wi: service weight. Number of cells to be served in T's for output queue i.

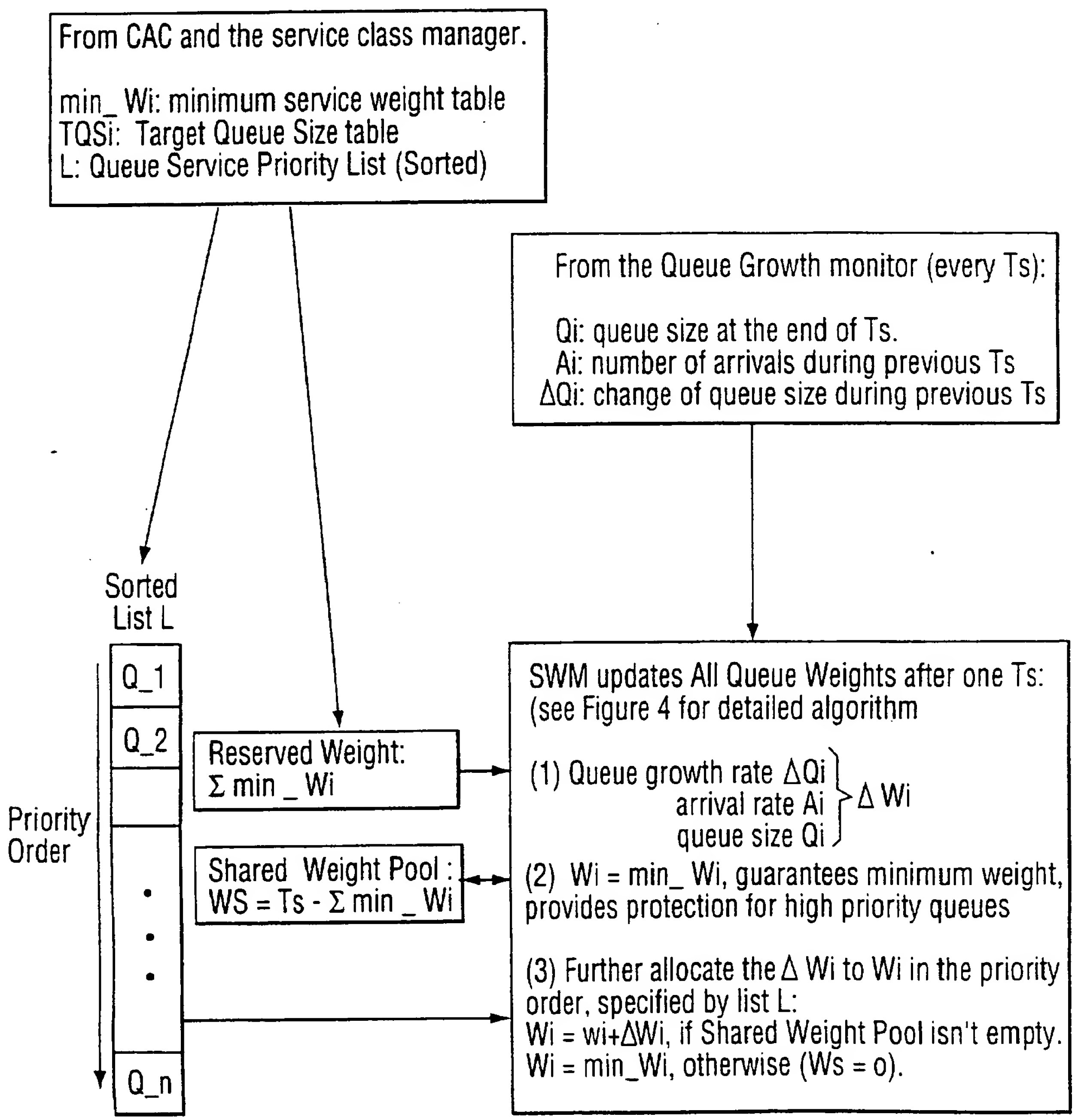


FIG. 3

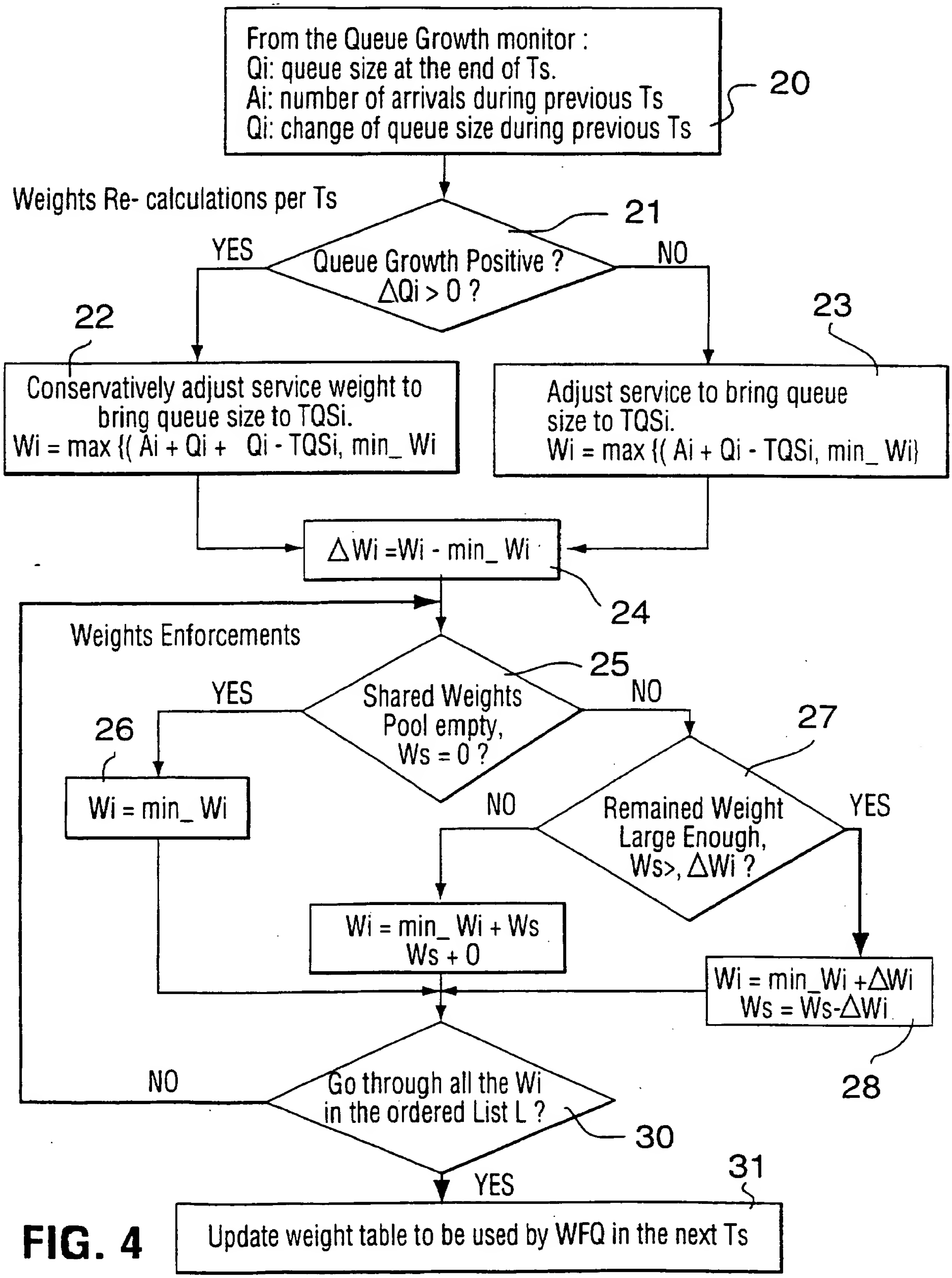

From the Queue Growth monitor :
Qi: queue size at the end of Ts.
Ai: number of arrivals during previous Ts
Qi: change of queue size during previous Ts
20
Weights Re- calculations per Ts
21
YES
Queue Growth Positive ?
ΔQi > 0 ?
NO
22
23
Conservatively adjust service weight to bring queue size to TQSi.
Wi = max {(Ai + Qi + Qi - TQSi, min_ Wi
Adjust service to bring queue size to TQSi.
Wi = max {(Ai + Qi - TQSi, min_ Wi}
ΔWi =Wi - min_ Wi
24
Weights Enforcements
25
YES
Shared Weights Pool empty, Ws = 0 ?
NO
26
27
Wi = min_ Wi
NO
Remained Weight Large Enough, Ws>, ΔWi ?
YES
Wi = min_ Wi + Ws
Ws + 0
Wi = min_Wi +ΔWi
Ws = Ws-ΔWi
28
NO
Go through all the Wi in the ordered List L ?
30
YES
31
Update weight table to be used by WFQ in the next Ts

FIG. 4

INTERNATIONAL SEARCH REPORT	Inter nal Application No PCT/CA 96/00681

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04L12/56 H04Q11/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04L H04Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 432 790 A (HLUCHYJ MICHAEL G ET AL) 11 July 1995 * Figures 4,5,6 * see column 1, line 56 - column 2, line 3 see column 3, line 3-25 see column 3, line 50-68 see column 4, line 41-54	1-5,7-11
A		6

A	EP 0 592 027 A (NEDERLAND PTT) 13 April 1994 see page 2, line 3-11 ---	4,5,9-11
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
20 January 1997	24. 01. 97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Authorized officer Dhondt, E

INTERNATIONAL SEARCH REPORT

Inter onal Application No
PCı/CA 96/00681

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTIN SYSTEMS, POZNAN, POLAND, JUNE 21 - 24, 1994, no. CONF. 14, 21 June 1994, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 268-275, XP000489090 GUHA A ET AL: "REAL-TIME SUPPORT OF CONTINUOUS AND VARIABLE BIT RATE TRAFFIC ON AN ATM NETWORK" * Figure 1 * see page 270, column 2, line 35-37 see page 270, column 2, line 45 - page 271, column 1, line 6 see page 271, column 1, line 30-43 -----	2,3,8

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/CA 96/00681

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US-A-5432790	11-07-95	AU-A-	7558294	03-04-95
		EP-A-	0680679	08-11-95
		WO-A-	9508230	23-03-95
EP-A-0592027	13-04-94	NL-A-	9201668	18-04-94
		JP-A-	6268660	22-09-94